



Laboratoire de Chimie de Coordination
CNRS, UPR 8241

CIF
Crystallographic Information File

Jean-Claude Daran
Directeur de recherche émérite CNRS

CIF

Cystallographic **I**nformation **F**ile

- *archiver les données cristallographiques*
- *transmettre ces données entre différents laboratoires*
- *transférer ces données d'un programme à un autre*
- *utiliser comme matériel supplémentaire*
- *soumission électronique à une revue scientifique*
- *transférer les résultats dans une base de données*

Un fichier **CIF** est toujours en **ASCII**.
Seul les caractères suivants sont permis

abcdefghijklmnopqrstuvwxyz

ABCDEFGHIJKLMNOPQRSTUVWXYZ

0123456789 !@#\$%^&*()_+ "~<>?|\-=[];'"/.

Tous les autres caractères tels que

- ◆ Å, °, é, Ø
- ◆ indices et exposants
- ◆ Lettres grecques
- ◆ ±, ≥ or ∞

nécessitent l'utilisation d'un code spécial

Ex. \%A pour Å ; \p pour π

Les unités sont implicites:

- Å pour `_cell_length_a`
- min pour `_diffrn_standards_interval_time`

Il faut donc utiliser `_cell_volume` 2367.5(8)
et non `_cell_volume` 2367.5(8)\%A^3^

L'utilisation d'un éditeur de texte ou
d'un éditeur spécial **CIF**
est fortement recommandé

Toutefois un traitement de texte (Word) peut-être
utilisé, mais le fichier créé en sortie doit être:

- au format **ASCII**
- ne pas contenir de codes invisibles ou cachés
- pas de ligne supérieure à **80 caractères**

Terminologie du CIF

Data name une ligne de caractères débutant par un underscore (_)

Data item une ligne de texte ne débutant pas par (_) mais qui est précédé par un **data name**.

Text string une série de caractères délimités par des espaces, des quotes, ou des points virgules comme premier caractère de la ligne.

`_exptl_crystal_density_method` 'not measured'

`_computing_molecular_graphics`

'ORTEPIII (Burnett & Johnson, 1996); ORTEP-3 for Windows (Farrugia, 1997)'

Terminologie du CIF

Data loop une liste de **data names**, précédé par `_loop` et suivie par une liste répétitive de **data items**

loop_

`_atom_type_symbol`

`_atom_type_description`

`_atom_type_scatter_dispersion_real`

`_atom_type_scatter_dispersion_imag`

`_atom_type_scatter_source`

`'C' 'C' 0.0033 0.0016`

`'International Tables Vol C Tables 4.2.6.8 and 6.1.1.4'`

`'H' 'H' 0.0000 0.0000`

`'International Tables Vol C Tables 4.2.6.8 and 6.1.1.4'`

`'O' 'O' 0.0106 0.0060`

`'International Tables Vol C Tables 4.2.6.8 and 6.1.1.4'`

Terminologie du CIF

Data block un ensemble de **data names** et de **data items** (qui peuvent être arrangés en boucle) précédé par une instruction `data_code` et terminé par une autre instruction `data_` ou par la fin du fichier.

data_Global

`_publ_contact_author_name` 'Daran Jean-Claude'

`_publ_contact_author_address`

;

Laboratoire de Chimie de Coordination,
205, route de Narbonne
31077 Toulouse Cedex 04, France

;

Data_compound1

Terminologie du CIF

Data file

un ensemble de **data blocks**: aucun **data blocks** ne doit avoir le même nom.

data_global

Data_compound1

Data_compound2

Data names et data blocks

- définitions jusqu'à 76 caractères
- assemblage hiérarchisé

`_<category>_<topic>_<subtopic`

- différentes possibilités de catégories comme

`_publ_ _exptl_ _geom_ _refln_`

- aisément compréhensible

e.g., `_publ_author_name`
`_exptl_crystal_colour`
`_computing_structure_solution`

■ **data blocks :**

(a) **data_global**

- modèle
- information concernant l'auteur à contacter
- information sur la soumission
- **liste des auteurs (nom, affiliation, adresse)**
- titre • résumé • commentaires
- liste des références
- Légendes des figures et des tableaux
- remerciements
- etc

(b) un **data block** pour chaque structure

(e.g., data_compound1, data_compound2,...)

Quelques propriétés du format CIF

- (a) L'ordre des couples **data name/data item** et l'ordre des **data blocks** n'a pas d'importance.
- (b) A chaque **data name** doit être associé un **data item**, ce dernier peut ne contenir aucune donnée réelle mais doit avoir un symbole tel que "?" ou ".", par exemple

```
_chemical_name_common      ?
```

Dans la boucle ci-dessous, le quatrième **data name** s'applique seulement dans la seconde ligne.

```
Loop_
_geom_bond_atom_site_label_1
_geom_bond_atom_site_label_2
_geom_bond_distance
_geom_bond_site_symmetry_2
_geom_bond_publ_flag
Ni      N1      2.036(2) . Yes
Ni      N1      2.054(2) 2_555 Yes
Ni      S2      2.421(10) . Yes
C       S2      1.637(3) . Yes
N1      C1      1.327(3) . ?
```

Attention, dans certains cas, certaines données sont obligatoires: e.g., pour Acta Cryst. C par exemple

- (c) **Les codes standards doivent être utilisés autant que possible**
e.g., traitement des atomes H durant l'affinement

```
_refine_ls_hydrogen_treatment          refall
```

ou

```
_refine_ls_hydrogen_treatment          constr
```

Sinon, il faut fournir une explication la plus complète possible.

- (d) **possibilité de préparer des tableaux**
standard, par exemple présentation des paramètres de liaisons H

```
_geom_hbond_<subtopic>
```

non-standard : il faut alors définir des données supplémentaires
e.g., comparaison de paramètres géométriques:

```
loop_
```

```
_publ_manuscript_incl_extra_item
```

```
'_geom_extra_tableA_col_1'
```

```
'_geom_extra_tableA_col_2'
```

```
'_geom_extra_tableA_col_3'
```

```
'_geom_extra_tableA_col_4'
```

```
'_geom_extra_tableA_col_5'
```

```
'_geom_extra_tableA_col_6' # jusqu'à 14 colonnes
```

```
'_geom_extra_table_head_A' # pour en tête de tableau
```

```
'_geom_table_headnote_A' # si nécessaire
```

```
'_geom_table_footnote_A' # si nécessaire
```

```

geom_extra_table_head_A # entête du tableau
;
Table 3.
Comparison of molecular geometry parameters (\%A,\%) for
1,3-dioxolan-2-ones
;

loop_
_geom_extra_tableA_col_1
_geom_extra_tableA_col_2
_geom_extra_tableA_col_3
_geom_extra_tableA_col_4
_geom_extra_tableA_col_5
_geom_extra_tableA_col_6

Parameter^a^ (I) (II) (III) (IV) (V)
O1---C2 1.33 1.327(2) 1.316(6) 1.34(2) 1.323(5)
"C2\ldb O2" 1.15 1.207(2) 1.192(6) 1.21(2) 1.200(6)
C2---O3 1.33 1.341(2) 1.316(6) 1.28(2) 1.348(6)
O3---C4 1.40 1.447(2) 1.443(5) 1.42(2) 1.460(6)
C4---C5 1.52 1.531(2) 1.498(7) 1.53(2) 1.527(6)
O1---C5 1.40 1.448(2) 1.420(6) 1.46(2) 1.456(5)
O1---C2---O3 111 112.7(1) 111.9(4) 113(1) 112.0(4)

_geom_table_footnote_A # légende du pied de page
;
(I) 1,3-dioxolan-2-one (Brown, 1954)
(II) D-erythronic acid 3,4-carbonate (Moen, 1982)
(III) 4-p-chlorophenyloxymethyl-1,3-dioxolan-2-one (Katzhendler et al.,1989)
(IV) 4-(5-(2-iodo-1-hydroxyethyl)-5-methyl-tetrahydro-2-furyl)-4-methyl-1,
3-dioxolan-2-one (Wuts, D'Costa & Butler, 1984)
(V) 1,6-bis(1,3-dioxolan-2-one)-2,5-dithiahexane (this work)

^a^ Numbering schemes have been standardised as for (V)
;

```

Quelques astuces pratiques

1. ligne de caractères – 3 façons de les utiliser

Une manipulation incorrecte conduit au chaos!!

(a) délimitée par des espaces

- pas d'espace dans la donnée (data item)
- ne peut pas excéder une ligne
- exemples:

```
_publ_contact_author_email      J.O'Groats@north.ac.uk
_cell_length_a                   10.446(3)
_diffrn_standards_number        3
```

FAUX:

```
J. O'Groats@north.ac.uk
10.446 (3)
```

(b) délimité par des apostrophes

- la donnée peut maintenant contenir des espaces
- elle reste limitée à une ligne
- mais cette ligne peut apparaître sur la ligne suivante

```
_exptl_crystal_density_method      'not measured'
_chemical_formula_moiety            'C12 H24 S6 Cu 2+, 2(P F6 -)'
_publ_section_acknowledgements
'We thank EPSRC for a postdoctoral award (to J.O'G.).'
```

(a) délimité par des points virgules pour les textes qui excèdent une ligne par exemple:

```
_publ_section_abstract
;
```

```
In the title compound C~12~H~22~O~7~, (1), molecules occur
exclusively as the cis geometric isomer and are linked by
hydrogen bonding to form infinite helices running parallel
to the crystallographic c direction.
```

```
;
```

2. Text

(a) Pour une publication (Acta Cryst.)

`_publ_section_abstract`

`_publ_section_comment`

`_publ_section_references`

- la plupart du temps, il s'agit de texte normal
- indices : un jeu de tildes `~~`
- exposants: un jeu de signe d'omission `^^`

e.g., pour $[\text{Cu}(\text{H}_2\text{O})_4]^{2+}$ il faut écrire `[Cu(H~2~O)~4~]^2+^`

- à éviter dans `_chemical_formula_moiety`, etc
- pour Å utiliser `\%A` • pour ° utiliser `\%`
- pour les autres codes - voir la référence 4

<http://journals.iucr.org/c/services/authorservices.html>

(b) Erreurs communes, triviales mais ou ennuyeuses

- Le logiciel fait ce que vous lui dites, mais pas ce que vous voulez
- défaut d'accord entre les codes indices et exposants
- défaut dans la fin d'une ligne ou d'un texte
- défaut dans l'utilisation correcte des points virgules
- Le logiciel de vérification fait ce qu'il peut, mais on peut avoir des erreurs ou des effets pervers pour des raisons mineures

(c) Selection de données géométriques

`_geom_type_publ_flag`

(*type est* bond, angle *ou* torsion)

- »» Yes (or y) indique que la valeur est sélectionnée
- »» en dehors de Y ou Yes, (No, n, ?) indique que la valeur n'est pas sélectionnée

Toujours être prudent en éditant un CIF - utiliser un éditeur spécial comme enCIFer réduira les risques d'erreurs. Une erreur fréquente est le retrait d'un point (.) ou d'un **espace**: ce qui changera le nombre de paramètres sur une ligne, conduisant à une confusion entre les **data names** et les **data items**.

Erreurs fréquentes

Les erreurs de syntaxes les plus fréquentes lors des soumission de fichiers CIF à Acta Crystallographica, Section C ou E sont:

`_publ_author_footnote`

Ce **data name** est stipulée dans la boucle mais aucune donnée n'est fournie. (Si le nombre d'auteurs est un multiple de 3, **checkCIF** ne détecte pas cette erreur).

`_publ_author_address`

oubli du point virgule (;) avant ou après l'adresse

Tableau de liaisons hydrogène

pas d' 'underscore' (_) dans le code de symétrie
(**2555** au lieu de **2_555**)

Si aucune sortie ne peut-être obtenue avec **printCIF** (voir ci-dessous) c'est souvent parce que persiste des erreurs de syntaxe dans le CIF. En particulier, vérifier que les symbols `~` et `^` pour les indices et les exposants vont bien par paires.

Vérification du CIF

■ envoyer le fichier CIF par e-mail à:

checkcif@iucr.org

ou utiliser le site web à l'adresse suivante:

<http://journals.iucr.org/services/cif/checking/checkform.html>

pour différents tests:

- validation des data names
- vérifier les erreurs de syntaxe
- vérifier la consistance des données cristallographiques
- vérifier le groupe d'espace (les symétries oubliées!)
- détecter les valeurs anormales des facteurs de température
- détecter des incompatibilités entre coordonnées et géométrie
- détecter des oublis (données indispensables pour Acta Cryst.)
- et bien d'autres choses

Certains logiciels incluent ces vérifications, soit par lien direct avec le site **IUCR**, soit parce que les programmes de vérifications sont inclus directement dans le logiciel.

Impression du fichier CIF

On peut aussi obtenir une impression au format d'une publication du fichier CIF (Acta Cryst.)

■ envoyer le fichier par e-mail à:

printcif@iucr.org

ou utiliser le web à l'adresse suivante:

<http://journals.iucr.org/services/cif/checking/printform.html>

Dans ce dernier, on peut obtenir un 'preprint' au format postscript ou PDF. Il faut bien sur vérifier qu'il ne contient pas d'erreurs.

Références

1. S.R. Hall, F.H. Allen & I.D. Brown, *Acta Cryst.* 1991, A47, 655–685.
Des copies sont disponibles auprès de
International Union of Crystallography,
5 Abbey Square, Chester CH1 2HU, England.
2. S.R. Hall, *J. Chem. Inf. Comput. Sci.*, 1991, 31, 326–333.
3. *Acta Cryst.* 2002, C58, e2–e8.
4. *A Guide to CIF for Authors.* Des copies sont disponibles auprès de
International Union of Crystallography,
5 Abbey Square, Chester CH1 2HU, England.

Alexander J. Blake, Co-Editor *Acta Crystallographica*
School of Chemistry, The University of Nottingham
University Park, Nottingham NG7 2RD, UK

Tel: (INT+) 44 115 951 3488 Fax: (INT+) 44 115 951 3563

<http://www.iucr.org/iucr-top/cif/software/>

enCIFer

Editeur graphique permettant:

- Localisation et indications des erreurs en utilisant le dictionnaire
- Correction des erreurs de syntaxe
- Edition de données individuelles ou en boucles
- Addition de nouvelles données individuelles ou en boucle
- Addition d'information standard grâce à des aides (*wizard*)
- Aide à la publication- information bibliographique demandée par la plupart des journaux
- 3D visualisation de la structure (**Mercury**)

Mercury

Visualisation de la molécule et calcul de géométrie à partir d'un fichier CIF mais aussi de a partir de différents formats.

PubCIF

A partir d'un fichier, prépare un fichier formaté (**Preprint**) dans le style *Acta Crystallographica Sections C et E*.

Le CIF et le fichier formaté (Preprint) sont présentés côte à côte et sont édités ensemble. Les corrections de l'un sont appliquées à l'autre.

Références logiciels

- **CIF** : <http://www.iucr.org/resources/cif/software/cif>
- **PublCIF** : <http://www.iucr.org/resources/cif/software/publCIF>
- **Mercury** : <http://www.ccdc.cam.ac.uk/products/mercury/>
- **PrintCIF**: <http://publCIF.iucr.org/services/tools/printCIF.php>
- **CheckCIF**: <http://checkCIF.iucr.org/>
- **PLATON** : <http://www.cryst.chem.uu.nl/platon/>
- **CIFTab** : <http://shelx.uni-ac.gwdg.de/SHELX/>